

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

**Enhancement Layer Switching
for Scalable Video Coding**

Inventor(s):
Jizheng Xu
Feng Wu
Shipeng Li

ATTORNEY'S DOCKET NO. MS1-1672US
CLIENT'S DOCKET NO. 305665.1

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25

4
5
6
7

8

9

10
11
12
13
14
15
16
17
18
19
20

21
22
23
24
25

1 necessary to achieve acceptable transfer speeds. In some cases, the loss of quality
2 may not even be detectable to the viewer.

3 Video compression is well known. One common type of video
4 compression is a motion-compensation-based video coding scheme, which is used
5 in such coding standards as MPEG-1, MPEG-2, MPEG-4, H.261, and H.263. In
6 such coding standards, video images are sampled and transformed into coefficients
7 that more or less capture the variation in pixels across the image. The coefficients
8 are then quantized and transmitted to a decoder. The decoder is able to decode the
9 image by performing operations that are substantially the inverse of the encoding
10 operations.

11 One particular type of motion-compensation-based video coding scheme is
12 fine-granularity layered coding. Layered coding is a family of signal
13 representation techniques in which the source information is partitioned into sets
14 called "layers". The layers are organized so that the lowest, or "base layer",
15 contains the minimum information for intelligibility. The base layer is typically
16 encoded to fit in the minimum channel bandwidth. The goal is to deliver and
17 decode at least the base layer to provide minimal quality video. The other layers,
18 called "enhancement layers", contain additional information that incrementally
19 improves the overall quality of the video. With layered coding, lower layers of
20 video data are often used to predict one or more higher layers of video data.

21 Another layered coding scheme is progressive FGS (PFGS). In PFGS, two
22 reference images are constructed for each frame, one is the reconstruction image of
23 the base layer, and the other is high quality reference image that is reconstructed
24 using the base layer bitstream and a part of the enhancement layer bitstream.
25 PFGS can improve coding efficiency over FGS because the prediction in PFGS is

1 based on higher quality enhancement layers, rather than only the low quality base
2 layer, as in FGS.

3 With layered coding, the various layers can be sent over the network as
4 separate sub-streams, where the quality level of the video increases as each sub-
5 stream is received and decoded. A decoder that receives the base layer and the
6 enhancement layers can be configured to choose and decode a particular subset of
7 these layers to get a particular quality according to its preference and capability.

8 Layered coding schemes are scalable, meaning that each layer can be scaled
9 in one or more aspects to achieve various desired performance goals. Spatial
10 scalability refers to approaches in which an image is decomposed into layers at
11 different spatial resolutions. Signal-to-noise (SNR) ratio scalability refers to
12 approaches in which the same spatial resolution is applied to the layers, but
13 coefficients are quantized at increasingly higher granularities.

14 While scalability can improve the visual quality of video, serious problems,
15 such as drifting, can occur. Drifting refers to a situation in which reference images
16 at the encoder and decoder do not match. In addition, coding efficiency can be
17 reduced when network bandwidth fluctuations are large. For example, when a
18 scalable video codec is set to optimize coding performance at a low bit-rate, often
19 the performance at high bit-rate will be sacrificed.

20 21 SUMMARY

22 Implementations of systems, methods, data structures, and computer
23 program products described herein solve the above, and other, problems and
24 drawbacks by providing a video encoding scheme that encodes video data frames
25 into multiple layers, including a base layer and at least two enhancement layers of

1 increasingly higher bit-rates, wherein only one of the enhancement layers is
2 transmitted based on a current network bandwidth and a switching bit-rate related
3 to the enhancement layer bit-rates.

4 An implementation of a system includes a data encoder generating a base
5 layer bitstream encoded at a base bit-rate, and a plurality of enhancement layer
6 bitstreams encoded at different enhancement layer bit-rates, and a bitstream
7 selection module selecting one of the enhancement layer bitstreams every video
8 frame based on available channel bandwidth.

9 An implementation of a method includes transmitting a first enhancement
10 layer bitstream encoded at a first bit-rate, detecting a transition in network
11 bandwidth through a switching bit-rate, and transmitting a second enhancement
12 layer bitstream encoded at a second bit-rate based on the transition in network
13 bandwidth.

14 An implementation of a computer program product provides a computer
15 program storage medium readable by a computer system and encoding a computer
16 program that dynamically switches among multiple enhancement layer bitstreams.
17 The implementation of the computer program product encodes a computer
18 program for executing on a computer system a computer process for dynamically
19 switching among multiple enhancement layer bitstreams. A base layer and
20 multiple enhancement layers are generated continually and a switching module
21 selects among the enhancement layers based on the bit-rates of the enhancement
22 layers and a current network bandwidth. The current network bandwidth is
23 analyzed with respect to one or more switching bit-rates associated with the
24 enhancement layer bit-rates to determine which enhancement layer to transmit.
25

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram of a video distribution system in which a content producer/provider encodes video data and transfers the encoded video data over a network to a client.

Fig. 2 is a block diagram of an exemplary video encoder without a difference bitstream generator implemented at the content producer/provider.

Fig. 3 is a diagrammatic illustration of an enhancement layer switching scheme without an associated difference bitstream, wherein switching is based on changes in available network bandwidth.

Fig. 4 is a block diagram of another exemplary video encoder implemented at the content producer/provider and including a difference bitstream encoder for reducing possible drifting errors caused by stream switching.

Fig. 5 is a diagrammatic illustration of an enhancement layer switching scheme with an associated difference bitstream, wherein switching is based on changes in available network bandwidth.

Fig. 6 is a flow diagram showing a method for determining one enhancement layer bitstream from a plurality of enhancement layer bitstreams based on network bandwidth.

DETAILED DESCRIPTION

This disclosure describes a layered video coding scheme used in motion-compensation-based video coding systems and methods. The coding scheme is described in the context of delivering video data over a network, such as the

1 Internet or a wireless network. However, the layered video coding scheme has
2 general applicability to a wide variety of environments.

3 Bandwidth fluctuation can be a major problem encountered in transmitting
4 scalable video over the Internet or wireless channels. The video coding scheme
5 described below can adapt to the channel conditions, while providing high coding
6 efficiency and performance.

8 **Exemplary System Architecture**

9 Fig. 1 shows a video distribution system 100 in which a content
10 producer/provider 102 produces and/or distributes video over a network 104 to a
11 client 106. The network 104 may represent of many different types of networks,
12 including the Internet, a LAN (local area network), a WAN (wide area network), a
13 SAN (storage area network), and wireless networks (e.g., satellite, cellular, RF,
14 etc.).

15 The content producer/provider 102 may be implemented in many ways,
16 including as one or more server computers configured to store, process, and
17 distribute video data. The content producer/provider 102 has a video storage 108
18 to store digital video files 110 and a distribution server 112 to encode the video
19 data and distribute it over the network 64. The server 112 has a processor 114, an
20 operating system 116 (e.g., Windows NT, Unix, etc.), and a video encoder 118.
21 The video encoder 118 may be implemented in software, firmware, and/or
22 hardware. The encoder is shown as a separate standalone module for discussion
23 purposes, but may be constructed as part of the processor 114 or incorporated into
24 operating system 116 or other applications (not shown).

1 The video encoder 118 encodes the video data 110 using a motion-
2 compensation-based coding scheme. One implementation of the video encoder
3 118 employs a progressive fine-granularity scalable (PFGS) layered coding
4 scheme. The video encoder 118 encodes the video into multiple layers, including a
5 base layer and one or more enhancement layers. The base layer represents an
6 encoded version of the video data 110 that contains the minimum amount of data
7 needed for decoding the video signal. Enhancement layers represent additional
8 information used to enhance the video signal representation when decoded at the
9 client 106. "Fine-granularity" coding means that the difference between any two
10 layers, even if small, can be used by a video decoder to improve the image quality.
11 Progressive Fine-granularity layered video coding ensures that the prediction of a
12 next video frame from a high quality reference of the current video frame is good
13 enough to keep the efficiency of the overall video coding.

14 In the illustrated implementation, the video encoder 118 includes a base
15 layer encoding component 120 to encode the video data 110 into the base layer, an
16 enhancement layer encoding component 122 to encode the video data into one or
17 more enhancement layers, an enhancement layer selection module 124 to select
18 one of the enhancement layers for transmission based on bandwidth of the network
19 104, and a difference layer encoding component 126. The video encoder 118
20 encodes the video data 110 such that each enhancement layer in a current frame is
21 independently based on the base layer. Exemplary implementations of the video
22 encoder 118 are described below in more detail with reference to Fig. 2 and Fig. 4.

23 Generally, the video data 110 is encoded on a frame-by-frame basis. In one
24 implementation, the base layer and the enhancement layers are concurrently
25 generated as video frames are input. Between two adjacent video frames, the

1 enhancement layer selection module 124 may switch from one enhancement layer
2 to another enhancement layer in response to changes in available network
3 bandwidth in order to improve data transmission performance. The enhancement
4 layer selection module 124 receives feedback from the network 104 regarding the
5 available network bandwidth. Many bandwidth detection techniques are known in
6 the art that may be employed by the enhancement layer selection module 124. By
7 way of example, a common way to measure bandwidth is to directly measure the
8 fastest rate that traffic can be sent through a network.

9 Based on the detected network bandwidth, the enhancement layer selection
10 module 124 chooses the enhancement layer that has a bit rate that is best fit for the
11 bandwidth. The term 'best fit' herein refers to the degree to which the selected
12 enhancement layer bit-rate matches a switching bit-rate determined by the
13 enhancement layer selection module 124. In one implementation, one or more
14 switching bit-rates are determined from a rate-distortion based function described
15 below with reference to Fig. 3. By selecting an enhancement layer that is best fit
16 for the current network bandwidth, various data transmission performance metrics
17 can be optimized, such as peak signal to noise ratio (PSNR).

18 In addition to dynamic switching among multiple enhancement layers, a
19 difference layer may be generated by a difference layer encoding component 126.
20 The difference layer encoding component 126 generates a frame-based difference
21 bitstream for determining the enhancement layer of a current frame with reference
22 to a different enhancement layer of a previous frame. The frame-based difference
23 bitstream represents the difference between the current enhancement layer and the
24 most recent enhancement layer switched from for reducing distortion between the
25 content provider/producer 102 and the client 106. An exemplary implementation of

1 the video encoder 118 that provides a frame-based difference bitstream is
2 described below in more detail with reference to Fig. 4.

3 The client 106 is equipped with a processor 128, a memory 130, and one or
4 more media output devices 132. The memory 130 stores an operating system 134
5 (e.g., a WINDOWS-brand operating system) that executes on the processor 128.
6 The operating system 134 implements a client-side video decoder 136 to decode
7 the layered video streams into the original video. In the event data is lost, the
8 decoder 136 is capable of reconstructing the missing portions of the video from
9 frames that are successfully transferred. Following decoding of each video frame,
10 the client 106 plays the video frame via the media output devices 132. The client
11 106 may be embodied in many different ways, including a computer, a handheld
12 entertainment device, a set-top box, a television, and so forth.

13 14 **Exemplary PFGS Layered Coding Scheme**

15 As noted above, the video encoder 118 encodes the video data into multiple
16 layers, such that a selected enhancement layer and a base layer in a current frame
17 are used to construct a high quality reference image for the frame. A parameter,
18 high quality reference bit-rate (HQRB) associated with each enhancement layer,
19 refers to the number of bits of the enhancement layer that are used to reconstruct
20 the high quality reference image. HQRB is discussed in more detail below. There
21 are many ways to implement a Progressive Fine-Granularity Scalable (PFGS)
22 layered video encoder to facilitate enhancement layer switching. One example is
23 illustrated in Fig. 2 for discussion purposes and to point out the advantages of the
24 scheme. Another example is illustrated below with a difference bitstream in Fig.
25 4.

Fig. 2 shows an exemplary implementation of a video encoder 200, which may be used by a server to encode video data files prior to distribution over a network. The video encoder 200 shown in Fig. 2 is based on the H.26L standard because of the standard's high coding performance. The video encoder 200 is configured to code video data using layered coding to facilitate an enhancement layer switching scheme, such as the scheme illustrated in Fig. 3 below.

Video encoder 200 has a base layer encoder 202 and an enhancement layer encoder 204, which are delineated by bold dashed boxes. The base layer encoder 202 produces a base layer bitstream 206 associated with a frame of video data. The enhancement layer encoder 204 includes two or more enhancement layer stages 208(1), 208(2), and so on, up to 208(n), which are separated by dashed-dotted lines. The enhancement layer stages 208(1), 208(2), up to 208(n), produce an enhancement layer bitstreams associated with each enhancement layer for each video frame.

Each of the stages 208(1), 208(2), up to 208(n) employs a feedback loop to track a high quality reference image and generate enhancement layer bitstreams 210(1), 210(2), up to 210(n), respectively, wherein 'n' is equal or greater than 2. As shown in Fig. 2, the enhancement layer bitstreams 210(1), 210(2), up to 210(n) may be generated simultaneously. The enhancement layer loops are independent from each other. As such, an i th enhancement layer stage 208(i), $1 \leq i \leq n$, reconstructs a high quality reference image using the base layer bitstream 206 and b_i bits of the enhancement layer bitstream 210(i) generated by the enhancement layer stage 208(i). Thus, the $HQRB_i$ for enhancement layer 210(i) is $b_i \times F$, where F denotes the frame rate.

1 After the base layer bitstream 206 and the enhancement layer bitstreams
2 210(i) are generated by the video encoder 200, a selection module, such as the
3 enhancement layer selection module 124 (Fig. 1), selects one of the enhancement
4 layer bitstreams 210 for transmission over the network. Selection of an
5 enhancement layer bitstream 210(i) is discussed below in more detail with regard
6 to Fig. 3.

7 With regard to the base layer encoder 202, video data is input frame-by-
8 frame to a motion estimator (ME) 212 to estimate the movement of objects from
9 locations in the one video frame to other locations in subsequent video frames. The
10 ME 212 also receives as reference for the current input, a base layer reference
11 image 219 stored in frame buffer 0 (214(0)) as well as one or more high-quality
12 reference images from the enhancement layers. The high-quality reference image
13 has a different signal-to-noise ratio (SNR) resolution than the based layer reference
14 219.

15 As illustrated in Fig. 2, the ME 212 uses the high-quality reference image
16 218 from the frame buffer 214(n) at enhancement layer stage 208(n) and the base
17 layer reference image 219. By using both the base layer reference image 219 and
18 the high quality reference image 218 for comparison with the input video to
19 determine motion vector(s), a good trade-off can be obtained between the base
20 layer 206 performance and the performance of the enhancement layers 210(i). By
21 using both the enhancement layer 210(n) reference and the base layer 206
22 reference in motion estimation, a significant gain can be achieved at a relatively
23 high network bandwidth bit-rate while maintaining almost the same performance
24 at a low network bandwidth bit-rate.
25

1 The ME 212 produces motion vectors that are sent to a motion compensator
2 (MC) 216(0). In a PFGS encoding scheme, both the base layer bitstream 206 and
3 the enhancement layer bitstreams 210(i) share the same set of motion vectors. The
4 output of the MC 216(0) is "RefB" signal, which represents the base layer
5 prediction. The RefB signal is input to loops in the enhancement layer encoder
6 204. The RefB signal is the result of applying motion displacement to a base layer
7 reference 219. The ME 212 and MC 214(0), MC 214(1), up to MC 214(n) are
8 well-known components used in conventional MPEG encoding.

9 An intra prediction generator 220 generates an intraframe (I-frame) block
10 prediction associated with input video data, where the prediction is generated from
11 the available information of neighboring blocks in the current frame, instead of
12 using the information from the previous frame.

13 In base layer coding performed by the base layer encoder 202, a frame
14 difference between the current input and the base layer prediction RefB signal is
15 divided into 8x8 blocks. Each residual block is transformed by a DCT (Discrete
16 Cosine Transform) module 222(0) and then quantized by a quantization (Q)
17 module 224. The base layer bitstream 206 is generated by coding the quantized
18 DCT coefficients using a variable length coder (VLC) 226(0). The output of the
19 VLC 226(0) is the base layer bitstream 206, which is transmitted to the decoder.

20 The base layers of the frames are also passed through an anti-quantization
21 (Q^{-1}) module 228. The de-quantized DCT coefficients are passed through inverse
22 DCT (IDCT) module 230(0). The output of the IDCT module 230(0) is added to
23 RefB in the base layer encoder 202. The output of the IDCT module 230(0), i.e.
24 the coded residues at the base layer in the current frame, is also fed into the
25

1 enhancement layer encoder 204, which is added to the temporal prediction of
2 enhancement layers 210(1), 210(2), up to 210(n).

3 The sum of output of the IDCT module 230(0) and the RefB signal is
4 filtered by loop filter 232(0). In one implementation, the loop filters 232 filter the
5 input signal according to the H.264 standard. The output of the loop filter 232(0)
6 is fed into the frame buffer 214(0). Thus, the frame buffer 214(0) represents the
7 decoded base layer bitstream 206 of the previous frame.

8 With regard to the enhancement layer encoder 204, the enhancement layers
9 are each independent from the other enhancement layers, and each enhancement
10 layer bitstream is generated using outputs from the base layer encoder 202.
11 Turning to the enhancement layer stages 208(1), 208(2), ..., 208(n), four signals
12 are input to each of the stages 208(1), 208(2), ..., 208(n): the coded residues
13 obtained from the IDCT module 230(0) in the base layer encoder 202, the motion
14 vectors from the motion estimator 212 in the base layer encoder 202, and the RefB
15 signal from the base layer encoder 202.

16 The enhancement layer encoder 204 receives the video data, the motion
17 vectors from the motion estimator 212, RefB from the motion compensator 216(0),
18 and the coded residues from IDCT module 230(0) and produces enhancement layer
19 bitstreams 210(1), 210(2), and so on, up to 210(n). At each stage 208(1), 208(2),
20 ..., 208(n) of the enhancement layer encoder 204, a motion compensator (MC)
21 216(0), 216(2), ..., 216(n) receives the motion vectors (MVs) from the base layer
22 encoder 202. Each MC 216(0), 216(2), ..., 216(n) also receives the output of a
23 frame buffer 214(0), 214(2), ..., 214(n) that stores the decoded enhancement layer
24 data from the previous frame. The MCs 216(0), 216(2), ..., 216(n) output motion
25 compensated data for the enhancement layer for the current frame. Either the

1 output from one of the MCs 216(1), 216(2), ..., 216(n) at each stage 208, or RefB
2 from the motion compensator 216(0) is added to the coded residues from the IDCT
3 module 230(0) from the base layer encoder 202, to form the enhancement layer
4 prediction.

5 The video data is subtracted from the enhancement layer prediction at each
6 enhancement layer stage. The result of the subtractions are respective displaced
7 frame difference (DFD) signals of the corresponding stage. The DFD signals in
8 stages 208(1), 208(2),...,208(n) are input into DCT modules 222(1), 222(2), ...,
9 222(n), respectively. The DCT modules 222(1), 222(2), ..., 222(n) generate DCT
10 coefficients.

11 The outputs of the DCT modules 222(1), 222(2), ..., 222(n) are input into
12 bit plane modules 234(1), 234(2), ..., 234(n), respectively. The bit plane modules
13 234(1), 234(2), ..., 234(n), represent each of the DCT coefficients as a binary
14 number of several bits, which are arranged in a zigzag order into a one-
15 dimensional array, referred to as a bit plane. The bit planes are stored in memory.
16 Here, there are n bit plane modules 234(1), 234(2), ..., 234(n), that encode n
17 enhancement layers that are subsequently coded by variable length coders (VLC)
18 226(1), 226(2), ...,226(n), respectively. The output of each VLC 226(1), 226(2),
19 ..., 226(n) is an enhancement layer bitstream 210(1), 210(2), ...,210(n) for the
20 respective stage 208(1), 208(2), ..., 208(n).

21 To track the enhancement layers, the output bit planes from the bit plane
22 modules 234(1), 234(2), ..., 234(n) are input into IDCT modules 230(1), 230(2),
23 ..., 230(n), respectively. The IDCT modules 230(1), 230(2), ..., 230(n) output
24 decompressed representations of the enhancement layers. The decompressed
25 enhancement layers from the IDCT modules 230(1), 230(2), ..., 230(n) are

1 summed with the enhancement layer prediction, and then filtered by loop filters
2 232(1), 232(2), ..., 232(n). Frame buffers 214(1), 214(2), ..., 214(n) receive and
3 store the respective decompressed, filtered enhancement layer data. The contents
4 of the frame buffers 214(1), 214(2), ..., 214(n) are fed back to the motion
5 compensators 216(1), 216(2), ..., 216(n), respectively.

6 With regard to the exemplary encoder 200 shown in Fig. 2, as illustrated,
7 only one enhancement layer's high quality reference, namely the high quality
8 reference 218 from enhancement layer stage 208(n) (the highest bit-rate
9 enhancement layer), is used to determine the motion vectors from the motion
10 estimator 212. In this particular implementation, other enhancement layers are not
11 involved in the motion estimation procedure. By using only one high quality
12 reference in this way, encoder complexity can be reduced.

13 Other implementations of the encoder 200 can use more than one high
14 quality reference from other enhancement layer stages for the motion estimation
15 procedure to meet design and/or performance criteria specific to a particular
16 implementation. In such implementations, motion estimation may use one or more
17 of the high quality references obtained from one or more of the frame buffers
18 214(1), 214(2), ..., 214(n).

19 As shown in Fig. 2, the number of bits used for the high quality reference
20 associated with each enhancement layer stage 208(i) is shown by b_i . In this
21 particular implementation of the encoder 200, the number of bits, b_i , is
22 successively higher for each higher numbered enhancement layer stage 208(i).
23 That is, $b_1 < b_2 < \dots < b_n$. This implies that the bit-rates, $HQRB_i$, associated with
24 each of the high quality reference images are $HQRB_1 < HQRB_2 < \dots < HQRB_n$.
25 Accordingly, for bitstream i ($1 \leq i \leq n$), the encoder only needs to generate bits from

1 bit-rate 0 to bit-rate $HQRB_n$, i.e. for each frame, from 0 bits to b_n bits. Such a
2 method can significantly reduce operations for bit-plane scanning and variable
3 length coding. This means that for enhancement layer bitstream 210(i), bits beyond
4 $HQRB_i$ are useless and do not need to be generated by enhancement layer stage
5 208(i) to reconstruct the enhancement layer reference; however bits beyond
6 $HQRB_i$ can improve quality at the decoder if they are transmitted. Using the
7 encoder shown in Fig. 2, multiple enhancement layer bitstreams can be generated
8 simultaneously in a low complexity encoder.

9 Fig. 3 illustrates an exemplary bitstream switching process, in which two
10 enhancement layer bitstreams are generated, but only one enhancement layer
11 bitstream is transmitted based on available network bandwidth and the $HQRB_i$
12 associated with the enhancement layer bitstreams. The bitstream switching
13 process depicted in Fig. 3 can be performed by an enhancement layer selection
14 module in conjunction with a base layer encoder and an enhancement layer
15 encoder.

16 A group of Progressive Fine-Granularity Scalable (PFGS) bitstreams 302 is
17 produced including base layer bitstreams 304, first enhancement layer bitstreams
18 306, and second enhancement layer bitstreams 308. The bitstreams are generated
19 frame by frame as video data is received. In each frame, all of the bitstreams in the
20 group 302 are concurrently generated. For clarity, each enhancement layer
21 bitstream of each frame is labeled according to frame number and enhancement
22 layer number. Thus, for example, the first enhancement layer bitstream 314 in
23 frame 1 is labeled EL1F1, the first enhancement layer bitstream 316 in frame 2 is
24 labeled EL1F2, and so on. Although Fig. 3 shows bitstreams corresponding to
25 only two enhancement layers, it is to be understood that bitstreams for more than

1 two enhancement layers may be generated and the enhancement layer selection
2 module 124 may switch among any number of enhancement layers that are
3 generated.

4 The base layer bitstreams 304 are composed of a continuous series of bits
5 representing encoded input video data. The base layer bitstreams 304 represent the
6 video at a relatively low resolution that is minimally acceptable for viewing. The
7 base layer bitstreams 304 are encoded at a base bit-rate, denoted as R_B . Bit-rate
8 refers to the number of bits used to represent a layer of a video frame. Bit-rate is
9 related to the amount of compression and quantization applied to the video data at
10 each layer. As an image is compressed more and/or quantized less, bit-rate is
11 reduced.

12 The first enhancement layer bitstreams 306 and the second enhancement
13 layer bitstreams 308 are encoded to achieve successively higher bit-rates in
14 corresponding reference image. As discussed above, the bit-rate of a reconstructed
15 reference image associated with an enhancement layer is referred to as High
16 Quality Reference Bit Rates (HQRBs). The term $HQRB_i$ means that the i th that b_i
17 bits in the i th enhancement layer bitstream are used to reconstruct a high quality
18 reference video frame. The first enhancement layer bitstreams 306 are encoded
19 with b_1 bits to achieve a reference image bit-rate of $HQRB_1$ and the second
20 enhancement layer bitstreams 308 are encoded with b_2 bits to achieve bit-rate
21 $HQRB_2$. The successively larger shaded rectangular areas in Fig. 3 are used to
22 illustrate that b_2 is greater than b_1 .

23 Thus, the first enhancement layer bitstreams 306 and the second
24 enhancement layer bitstreams 308 satisfy $HQRB_1 < HQRB_2$. At a network
25 bandwidth bit-rate of $R_B + HQRB_2$, the second enhancement layer bitstream 308

1 will generally outperform the first enhancement layer bitstream 306, since the
2 second enhancement layer bitstream 308 has higher quality references. In
3 addition, drifting error that might have occurred in traditional codecs due to such
4 scaling will be reduced or avoided entirely due to a method, described below, for
5 choosing at what bit-rate to switch from the first enhancement layer bitstream 306
6 to the second enhancement layer bitstream 308.

7 Video data transmission performance can be improved by switching
8 between enhancement layer bitstreams of different HQRB. With regard to the first
9 enhancement layer bitstream 306 and the second enhancement layer bitstream 308,
10 the first enhancement layer bitstream 306 will typically outperform the second
11 enhancement layer bitstream 308 at comparatively low network bandwidth because
12 $HQRB_1$ is better matched to the low bit-rates; however, at comparatively higher
13 network bandwidth, the second enhancement layer bitstream 308 typically
14 performs better because $HQRB_2$ is better matched to the higher bit-rates.

15 Accordingly, when the network bandwidth exceeds a particular bit-rate
16 (called the switching bit-rate), the video encoding and transmission procedure will
17 switch from the first enhancement layer bitstream 306 to the second enhancement
18 layer bitstream 308. Switching from a lower HQRB bitstream to a higher HQRB
19 bitstream (e.g., from $HQRB_1$ to $HQRB_2$) is referred to as switching up. Switching
20 from a higher HQRB bitstream to a lower HQRB bitstream (e.g., from $HQRB_2$ to
21 $HQRB_1$) is referred to as switching down. The bitstream can be switched at each
22 frame.

23 With regard to the exemplary situation illustrated in Fig. 3, an enhancement
24 layer selection module dynamically switches between the first enhancement layer
25 bitstream 310 and the second enhancement layer bitstream 312 according to

1 changes in available network bandwidth. A bandwidth graph 310 illustrates how
2 available network bandwidth may change over time. At time t_0 , available
3 bandwidth starts at a low bandwidth, R_L . At time t_1 , the available bandwidth
4 increases to a high bandwidth, R_H . Later at time t_2 , the available bandwidth
5 decreases back to the low bandwidth, R_L . A switching bandwidth, R_S , exists
6 between the low bandwidth R_L and the high bandwidth R_H .

7 As time progresses, the base layer bitstream 304, the first enhancement
8 layer bitstream 306, and the second enhancement layer bitstream 308 are
9 concurrently generated at each frame as shown. Output bitstreams 312 are
10 generated frame by frame as the base layer bitstreams 304, the first enhancement
11 layer bitstreams 306, and the second enhancement layer bitstreams 308 are
12 generated. The base layer bitstream 304 is continuously transmitted, but only one
13 of the enhancement layer bitstreams is included in the output bitstream 312 in any
14 frame. At time t_0 , when the available bandwidth is R_L , the first enhancement layer
15 bitstream for frame 1 (EL1F1) 314 is selected. During frame 2, the bandwidth is
16 still R_L , so the first enhancement layer bitstream (EL1F2) 316 is selected.

17 When the bandwidth increases to R_H at time t_1 , the enhancement layer
18 selection module 124 detects a bandwidth transition through the switching
19 bandwidth, R_S . In response, a higher HQRB layer is switched to. Thus, at frame 3,
20 the enhancement layer 2 (EL2F3) 318 is selected and transmitted in the output
21 bitstream 312. At frame 4, the bandwidth remains above R_S at R_H , so the
22 enhancement layer bitstream (EL2F4) 320 is selected.

23 When the bandwidth decreases to R_L at time t_2 , the enhancement layer
24 selection module 124 detects a bandwidth transition through the switching
25 bandwidth, R_S and inserts the first enhancement layer bitstream (EL1F6) 322 into

1 the output bitstream 312. Thus, at each frame an enhancement layer bitstream is
2 selected and transmitted depending on the available network bandwidth and the
3 HQRBs of the first enhancement layer bitstreams 306 and the second enhancement
4 layer bitstreams 308.

5 The switching bit-rate may be determined for the first enhancement layer
6 bitstream 306 and the second enhancement layer bitstream 308 as follows.
7 Because $HQRB_1$ is less than $HQRB_2$, better performance is achieved by
8 transmitting the first enhancement layer bitstream 306 when the available network
9 bandwidth is in the range $[0, HQRB_1]$. Similarly, when the available network
10 bandwidth is equal to or greater than $HQRB_2$, better performance is achieved by
11 transmitting the second enhancement layer bitstream 308 because there is less
12 drifting error and the second enhancement layer bitstream 308 serves as a better
13 reference.

14 When the network bandwidth is between $(HQRB_1, HQRB_2)$, it is not easy to
15 tell which of the bitstreams should be transmitted. A rate-distortion based method
16 can be advantageously applied to estimate the switching bit-rate, R_S , which gives a
17 clear segmentation, wherein below R_S , the first enhancement layer bitstream 306 is
18 chosen and beyond R_S , the first enhancement layer bitstream 308 is switched to.
19 The estimation problem can be solved by estimating each bitstream's distortion at
20 a certain bit-rate. Suppose that an n-frame sequence is coded. For each frame,
21 since the image decoded is the sum of prediction image and residue signals, the
22 distortion of this frame consists of two sources. One source is the distortion of the
23 residue signals, and the other source is the distortion of the prediction image,
24 which caused by drifting error. For the entire sequence, the distortion includes
25 these two sources too, which can be characterized by the following equation:

$$(1) \quad D(R) = D_{res}(R) + D_{drifting}(R)$$

where $D_{res}(R)$ and $D_{drifting}(R)$ denote the distortion of residue images and drifting error respectively. $D_{res}(R)$ can be determined in a PFGS encoder, such as the encoder 118, because the residue images are coded bit-plane by bit-plane. In addition, if it is assumed that drifting error caused by each frame is independent, $D_{drifting}(R)$ can be estimated as:

$$(2) \quad D_{drifting}(R) = \sum_{i=1}^n D_{drifting}^i(R)$$

$D_{drifting}^i(R)$ is the drifting error caused by the frame i , $1 \leq i \leq n$. $D_{drifting}^i(R)$ is caused by the current frame's mismatch between the reference images of the encoder and the decoder. And this mismatch, denoted as $Diff^i(R)$ can be calculated by comparing the encoder's reference image and the decoder's reference at R that is also available in the encoder. Thus, if network bandwidth, R is greater than or equal to HQRB, $Diff^i(R)$ will be 0, as will $D_{drifting}(R)$. Note that not all macro-blocks suffer drifting error of the enhancement layer. For those macro-blocks referenced from the base layer, they are not affected by $Diff^i(R)$. Let c_j be the percentage of the macro-blocks of the frame j that use the high quality reference. Then the intensity of influence of $Diff^i(R)$ on the frame j can be estimated as $Diff^i(R) \prod_{k=i+1}^j C_k$, so $D_{drifting}^i(R)$ is estimated by:

$$(3) \quad D_{drifting}^i(R) = \sum_{j=i+1}^n (Diff^i(R) \prod_{k=i+1}^j C_k)$$

Equations (2) and (3) can be used to derive equation (4):

$$(4) \quad D_{drifting}(R) = \sum_{i=1}^n D_{drifting}^i(R) = \sum_{i=1}^n \sum_{j=i+1}^n (Diff^i(R) \prod_{k=i+1}^j C_k)$$

Accordingly, by summing $D_{res}(R)$ and $D_{drifting}(R)$, $D(R)$ can be estimated at a given bit-rate R .

For any 2 enhancement layer bitstreams, the switching bit-rate R_S is the bit-rate at which the first enhancement layer bitstream 306 and the second enhancement layer bitstream 308 have the same $D(R)$ value, denoted as

$$(5) \quad D_1(R_S) = D_2(R_S)$$

If $R < R_S$, then $D_1(R) < D_2(R)$. Conversely, when $R > R_S$, $D_1(R) > D_2(R)$. As mentioned above, R_S should be within $(HQRB_1, HQRB_2)$. Equation (5) can be solved by a bisection method wherein the initial interval is $(HQRB_1, HQRB_2)$. As known in the art, the bisection method proceeds by evaluating the function in question at the midpoint of the original interval $(HQRB_1 + HQRB_2)/2$ and testing to see in which of the subintervals $[HQRB_1, (HQRB_1 + HQRB_2)/2]$ or $[(HQRB_1 + HQRB_2)/2, HQRB_2]$ the solution lies. The estimation method can be extended to implementations that employ multiple enhancement layer bitstreams, in which a switching bit-rate is estimated for each pair of enhancement layer bitstreams.

An enhancement layer switching strategy can be implemented based on the above equations and results. An enhancement layer switching strategy refers to the policy of switching from one enhancement layer to another; i.e., under what conditions the decision is made to switch enhancement layers. An exemplary

1 strategy is based on the available network bandwidth or bit-rate. As mentioned
2 above, an estimate can be made as to which bitstream will work better given the
3 bitstream's bit-rate and the available network bandwidth. The estimate can be
4 improved by considering drifting that may be caused by switching.

5 If switching occurs, the high quality reference changes, and the decoder
6 does not have previous frames of the bitstream switched to, there may be a
7 reference mismatch. Thus, drifting error emerges. When bandwidth fluctuation is
8 high, drifting error can be exacerbated by frequent enhancement layer switching.
9 Performance under such conditions may be improved by making the encoder less
10 sensitive to fluctuations in network bandwidth. To improve the estimate of the
11 switching bit-rate, R_S , a buffer bit-rate, R_{Buf} , can be added to the original estimate
12 of R_S . The buffer bit-rate, R_{Buf} , can improve performance by reducing the number
13 of frequent enhancement layer switches. Mathematically, a new switching bit-rate,
14 R'_S , can be expressed as follows:

$$(6) \quad R'_S > R_S + R_{Buf}$$

15
16
17
18 Fig. 4 is a block diagram of another exemplary video encoder 400
19 implemented at a content producer/provider and including a difference bitstream
20 encoder 402. As with the encoder 200 shown in Fig. 2, the video encoder 400 of
21 Fig. 4 includes a base layer encoder 404, a first enhancement layer stage 406(1),
22 and second enhancement layer stage 406(2). The base layer encoder 404, the first
23 enhancement layer encoder 406(1), and second enhancement layer encoder 406(2),
24 carry out operations analogous to the base layer encoder 202, the first enhancement
25 layer stage 208(1), and the second enhancement layer stage 208(2), respectively,

1 which are shown and describe above with respect to Fig. 2. Therefore, the base
2 layer encoder 404, the first enhancement layer encoder 406(1), and the second
3 enhancement layer encoder 406(2) are not discussed in detail here.

4 The main relevant difference between the video encoder 400 and the video
5 encoder 200 shown in Fig. 2, is the difference bitstream encoder 402. In general,
6 the difference bitstream encoder 402 receives and combines data from the first
7 enhancement layer stage 406(1) and data from the second enhancement layer stage
8 406(2), and encodes the combination of the data to form a difference bitstream
9 408. The difference bitstream represents a difference between a high quality
10 reference image 410 from second enhancement layer stage 406(2) and the
11 enhancement layer prediction from the first enhancement layer stage 406(1).

12 The high quality reference image 410 is subtracted from the enhancement
13 layer prediction of the enhancement layer 412 from the first enhancement layer
14 stage 406(1) to create a difference signal 414. The difference signal is transformed
15 by a discrete cosine transform (DCT) function of DCT module 416. The DCT
16 module 416 outputs DCT coefficients that are encoded into a bit plane by a bit
17 plane module 418. The output of the bit plane module 416 is input to a variable
18 length coder (VLC) 420, which variable length encodes the bit plane. The output
19 of the VLC 420 is the enhancement layer difference bitstream 408.

20 Fig. 5 is a diagrammatic illustration of an exemplary difference bitstream
21 generation scheme 500 that may be employed in response to detection of network
22 bandwidth. The difference bitstream generation scheme 500 can be realized using
23 the video encoder 400 illustrated in Fig. 4. A group of PFGS video layer
24 bitstreams 502 are generated frame-by-frame. The group 502 includes base layer
25

1 bitstreams 504, first enhancement layer bitstreams 506, second enhancement layer
2 bitstreams 508 and enhancement layer difference bitstreams 510.

3 Output bitstreams 512 are output by the video encoder frame-by-frame. In
4 each frame, the output bitstream 512 includes a base layer bitstream from the base
5 layer bitstream 504, and one of either the first enhancement layer bitstreams 506
6 and the second enhancement layer bitstreams 508, or the difference bitstream 510.
7 In frames that the output bitstream 512 includes either the first enhancement layer
8 bitstream 506 or the second enhancement layer bitstream 508, the transmitted
9 enhancement layer bitstream is selected based on detected network bandwidth, in a
10 manner as is discussed above with respect to Fig. 3.

11 Thus, the exemplary situation illustrated in Fig. 5 includes a switch up
12 between frame 2 and frame 3. The exemplary situation assumes that the detected
13 network bandwidth transitioned from a low network bandwidth to a high network
14 bandwidth, and that the bandwidth transitioned through a switching bit-rate. Prior
15 to the bandwidth transition, a selected first enhancement layer bitstream 514 is
16 transmitted in the output bitstream 512.

17 After the bandwidth transition occurs, and in response to the bandwidth
18 transition, a scalable difference bitstream 516 is transmitted in the output bitstream
19 512 in frame 3, in place of an enhancement layer bitstream. Thus, when the
20 difference bitstream 516 is transmitted, the base layer bitstream 504 is transmitted
21 but the enhancement layer bitstream is not. As discussed earlier, the difference
22 bitstream 516 represents a difference between the first enhancement layer
23 bitstream 514 and a second selected enhancement layer bitstream 518. In frame 4,
24 the second selected enhancement layer bitstream 518 is transmitted.

1 Advantageously, the scalable difference bitstream 516 enables the
2 transmitting computer to flexibly decide how many bits should be dedicated to the
3 difference bitstream 516 when enhancement layer bitstream switching occurs.
4 Drifting between the server computer and the client computer can be greatly
5 reduced or eliminated if all the bits in the difference bitstream 516 are received by
6 the client. Thus, the encoder may truncate the difference bitstream 516 prior to
7 transmission (i.e., send fewer than all the difference bitstream bits). In addition,
8 no extra frame's bits need to be sent and decoded. As illustrated in Fig. 5, in the
9 bitstream switching process, including the difference bitstream 516 does not
10 impact the number of frames decoded. The decoder can decode the difference
11 bitstream 516 as a standard PFGS bitstream, which means that the bitstream
12 switching process with drifting error compensated is transparent to the decoder.

13 Fig. 6 is a flow diagram showing an enhancement layer bitstream switching
14 operation 600 for selecting one enhancement layer bitstream from a plurality of
15 enhancement layer bitstreams based on network bandwidth. The operation 600
16 may further output a difference bitstream when a switch is made from one
17 enhancement layer bitstream to another enhancement layer bitstream. The
18 operation 600 can be performed by the encoder 118, described in Fig. 1.

19 After a starting operation 602, a generating operation 604 receives a frame
20 of video data and generates a base layer bitstream, two or more enhancement layer
21 bitstreams, and, optionally, one or more difference bitstreams. The generating
22 operation 604 preferably, although not necessarily, generates the bitstreams
23 concurrently as a video frame is input.

24 A determining operation 604 determines one or more switching bit-rates
25 based on high quality reference bit rates (HQRBs) associated with enhancement

1 layer bitstreams. For each pair of enhancement layer bitstreams generated in the
2 generating operation 604, a switching bit-rate is determined between the two
3 enhancement layer bitstreams in the pair. As discussed above, a rate-distortion
4 based method can be employed to determine the one or more switching bit-rates.
5 In the rate-distortion based method, a switching bit-rate may be determined
6 between two HQRBs by using a bisection method discussed above.

7 A detecting operation 608 detects the current network band. Network
8 bandwidth may be detected using methods known in the art, such as network
9 probing and modeling. A selecting operation 610 selects one of the previously
10 generated enhancement layer bitstreams based on the HQRBs of the enhancement
11 layers, the one or more switching bit-rates, and the detected network bandwidth. In
12 one implementation, the selecting operation 610 first identifies the switching bit-
13 rate that is closest to the detected network bandwidth. Then the selecting operation
14 610 determines the HQRB of an enhancement layer that is closest to the switching
15 bit-rate. The enhancement layer with an HQRB closest to the identified switching
16 bit-rate is selected in the selecting operation 610.

17 An optional outputting operation 612 outputs a difference bitstream if the
18 selected enhancement layer for the current frame is different from the selected
19 enhancement layer of the previous frame. A second outputting operation 614
20 outputs the base layer bitstream and the selected enhancement layer bitstream.

21 Although some exemplary methods, devices and exemplary systems have
22 been illustrated in the accompanying Drawings and described in the foregoing
23 Detailed Description, it will be understood that the methods and systems are not
24 limited to the exemplary embodiments disclosed, but are capable of numerous
25 rearrangements, modifications and substitutions without departing from the spirit

1 set forth and defined by the following claims.

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25